In the name of Allah, the Most Gracious, the Most Merciful

LA FACULTÉ

SEEK KNOWLEDGE

Copyright disclaimer

"La faculté" is a website that collects copyrights-free medical documents for non-lucratif use.
Some articles are subject to the author's copyrights.
Our team does not own copyrights for some content we publish.
"La faculté" team tries to get a permission to publish any content; however , we are not able to contact all the authors.
If you are the author or copyrights owner of any kind of content on our website, please contact us on: facadm16@gmail.com
All users must know that "La faculté" team cannot be responsible anyway of any violation of the authors' copyrights.
Any lucrative use without permission of the copyrights' owner may expose the user to legal follow-up.

Partie 3:
Statistique Inférentielle

Chapitre I: Estimation

I/ Introduction:

Le problème de l'estimation stat. est le suivant:

On cherche à connaître les valeurs de certaines caractéristiques (paramètres) d'une variable aléatoire grâce à des observations réalisées sur un éch.

ex:

a/ Quelle est la fréquence d'une certaine maladie dans cette pop.?

b/ Quelle est la moyenne de la glycémie?

Il est impossible de répondre de façon sûre à ces questions. On y apporte 2 types de réponses:

- On produit une valeur qui nous semble être la «meilleure» possible: c'est l'estimation ponctuelle.
- On produit un intervalle de valeurs possibles, compatibles avec les observations: c'est la notion d'intervalle de confiance IC.

N.B.:

Dans toute la suite, la variable aléatoire d'intérêt est notée X, et le paramètre à estimer θ.

$m = E(X)$ = moyenne de la pop ou moyenne théorique.

$\sigma^2 = Var(X)$ = variance de la pop.

$P_0 = P(A)$ = proportion de A dans la pop.

Paramètres théoriques

$\bar{X}$ = moyenne de l'échantillon.

S^2 = variance de l'éch.

p = % de l'éch.

Paramètres calculés observés

II/ Estimation ponctuelle:

1/ Déf:

À partir d'un éch $(X_1, X_2, \dots, X_n)$ de la v. a. X, on construit une nouvelle v. a. $t(X_1, X_2, \dots, X_n)$ dont les réalisations se rapprochent de la valeur de θ.

Cette nouvelle variable, notée T ou $T_n = t(X_1, X_2, \dots, X_n)$ est appelée estimateur du paramètre θ

ex: $T = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$ est l'estimateur naturel de $m = E(X)$.

Un estimateur est une fonction de l'éch.

2/ Propriétés:

Un estimateur étant une fonction de l'éch. est donc, une v. a.

Donc, il possède une densité, une moyenne et une variance.

Ces deux grandeurs permettent de comparer, dans une certaine mesure, plusieurs estimateurs entre eux.

a/ Biais: d'un estimateur, noté $B(T)$ est la différence moyenne entre sa valeur et celle de la quantité qu'il estime.

$$B(T) = E(T-\theta) = E(T) - \theta.$$

Un estimateur T sans biais de θ (TESB de θ) ssi $B(T) = 0$

$$\Updownarrow$$

$$E(T) = \theta.$$

b/ Variance: d'un estimateur T est définie par: $E(T^2) - E^2(T) = V(T)$

$$E(T - E(T))^2$$

Si 2 estimateurs sont sans biais, le meilleur d'entre eux est celui qui a le plus petite variance.

c/ Erreur Quadratique Moyenne

EQM: elle permet de comparer plusieurs estimateurs entre eux qu'ils soient biaisés ou non.

$$EQM(T) = E[(T-\theta)^2]$$

$$= Var(T) + [B(T)]^2.$$

N.B.: si TESB de θ, $EQM(T) = Var(T)$.

Le meilleur estimateur est celui qui a la plus petite erreur quadratique moyenne.

3/ Exemples:

On montre que:

i/ $\bar{X}$ est "le meilleur" estimateur de m

$\theta = m$

On écrit: $\hat{m} = \bar{x}$

↓

valeur estimée de m

$\Updownarrow$

$m \simeq \bar{x}$

$m = \bar{x}$ Jamais

ii/ $\theta = \sigma^2$ On pourrait estimer σ^2 par s^2.

mais s^2 n'est pas un ESB de σ^2.

Alors, on lui préférera:

$$s'^2 = \frac{n}{n-1} s^2$$ qui est un ESB de σ^2

On écrira: $\hat{\sigma}^2 = \frac{n}{n-1} s^2 = s'^2$.

N.B.: $\frac{n}{n-1} s^2 = \frac{1}{n} \cdot \frac{n}{n-1} \sum (x_i - \bar{x})^2$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= s'^2.$$

s'^2 est appelée variance corrigée

iii/ $\theta = p_0$ p est « le meilleur » estimateur de p_0.

$\hat{p}_0 = p$.

On écrit:

$\bar{X}$ est une v.a., elle prend différentes valeurs selon l'éch. choisi d'où l'intérêt de ne pas s'arrêter à l'estimation ponctuelle, qui est reflet de l'éch., mais de donner un intervalle qui a de fortes chances de contenir le paramètre m

II/ Estimation par intervalle de confiance:

1/ Déf: estimer un paramètre θ par un intervalle de confiance (a, b) avec une certaine proba. $1-\alpha$, est équivalent à dire: $P(\theta \in (a,b)) = 1-\alpha$

Cette proba. est appelée le niveau de confiance / de signification / taux de sécurité, et α est appelé le risque ou le seuil ou l'erreur

N.B.:

α est fixé au préalable.

$\alpha \leq 10\%$ (ne doit jamais dépasser 10%). Souvent $\alpha = 5\%$.

1/ Intervalle de confiance pour la moyenne:

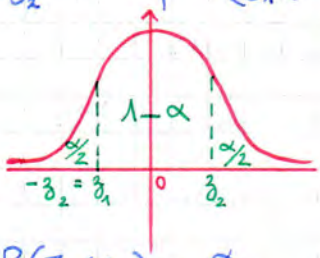
$\theta = m$ $a = ?$ $b = ?$ tq $P(a<m<b) = 1$.

a/ $n > 30$:

i/ σ^2 connu:

TCL $\Rightarrow \bar{X} \rightsquigarrow N\left(m, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-m}{\sigma/\sqrt{n}} \rightsquigarrow N(0,1)$.

$z_1 = ?$ $z_2 = ?$ tq $P(z_1 < Z < z_2) = 1-\alpha$.



z_1: $P(Z \leq z_1) = \frac{\alpha}{2}$. $z_1 = q_{\alpha/2}$

z_2: $P(Z \leq z_2) = 1-\frac{\alpha}{2}$ $z_2 = q_{1-\frac{\alpha}{2}}$

$$P\left(-q_{1-\frac{\alpha}{2}} < \frac{\bar{X}-m}{\sigma/\sqrt{n}} < q_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Updownarrow$$

$$P\left(\bar{X} - q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < m < \bar{X} + q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

On écrit: $IC_\alpha(m) = \left[\bar{X} \pm q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$

$= \left[\bar{X} - q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + q_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$

ii/ σ inconnu:

σ inconnu alors on l'estime.

$\hat{\sigma}^2 = \frac{n}{n-1}S^2 \Rightarrow \hat{\sigma} = \sqrt{\frac{n}{n-1}}\,S$

Soit $T = \frac{\bar{X}-m}{\hat{\sigma}/\sqrt{n}} = \frac{\bar{X}-m}{S/\sqrt{n-1}} \rightsquigarrow N(0,1)$

On obtient:

$$IC_\alpha(m) = \left[\bar{X} \pm q_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}}\right]$$

b/ $n \leqslant 30$

On supposera la Normalité de la pop.

i/ σ^2 connu

$$IC_\alpha(m) = \left[\bar{X} \pm q_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

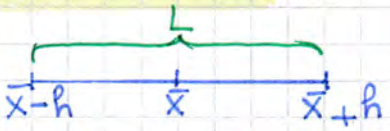
ii/ σ^2 inconnu

$$IC_\alpha(m) = \left[\bar{X} \pm t_{n-1}(\alpha) \frac{S}{\sqrt{n-1}}\right].$$

Précision d'un IC:

$IC_\alpha(m)$

$$IC_\alpha(m) = [\bar{X} \pm h].$$

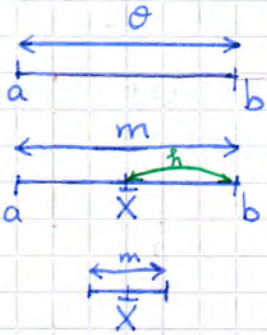


L'$IC_\alpha(m)$ est centré en $\bar{X}$.

$h = \frac{L}{2}$ = précision.

où L: longueur de l'intervalle

$L = b - a$; $h = \frac{b-a}{2}$.



Le meilleur IC et le plus précis est celui qui est le plus petit

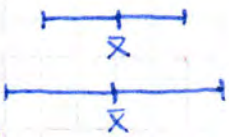
$m \in \mathbb{R}^+_*$ $\alpha = 0$

$$P(a < \theta < b) = 1 - \alpha.$$

Risque et précision:

$\alpha_1 = 5\%$ $IC_{\alpha_1}(m)$

$\alpha_2 = 1\%$ $IC_{\alpha_2}(m)$



$IC_{\alpha_1}(m) \subset IC_{\alpha_2}(m)$ $\forall \alpha_1 > \alpha_2$

Taille de l'éch.:

n?

n = ? $h = h_0 =$ cste.

- Trouver n en fixant la précision avec laquelle le paramètre a été estimé.

ex: $IC_\alpha(m) = [\bar{X} - h \; ; \; \bar{X} + h]$.

$n > 30$

σ^2 inconnu $h = q_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}} = h_0$.

$$n = \left(\frac{q_{1-\frac{\alpha}{2}} \, S}{h_0}\right)^2 + 1.$$

N.B.:

1- $n \leqslant 30$

σ^2 inconnu $h = t_{n-1}(\alpha) \frac{S}{\sqrt{n-1}} = h_0$

2- n = ?

Pour que la précision soit d'au moins h_0:

n = ? ~~$h \leqslant h_0$~~

n = ? $h \leqslant h_0$

Valeurs particulières de $q_{1-\frac{\alpha}{2}}$:

α	1%	5%	10%
$1-\alpha$	99%	95%	90%
$q_{1-\frac{\alpha}{2}}$	2,58%	1,96%	1,64%

$$IC_\alpha(m) = \bar{x} \pm h$$

Tableau donnant h:

	σ^2 connu	σ^2 inconnu
$n > 30$	$q_{1-\frac{\alpha}{2}} \sigma \frac{1}{\sqrt{n}}$	$q_{1-\frac{\alpha}{2}} S \frac{1}{\sqrt{n-1}}$
$n \leq 30$	$q_{1-\frac{\alpha}{2}} \sigma \frac{1}{\sqrt{n}}$	$t_{n-1}(\alpha) \frac{1}{\sqrt{n-1}}$

2/ Intervalle de confiance pour la proportion :

$p_0 = P(A)$ dans la pop.

$p = P(A)$ dans l'éch.

$a = ?$ $b = ?$ tq $P(a < p_0 < b) = 1-\alpha$

a/ $n > 30$:

P v.a. on montre $P \rightsquigarrow N\left(p_0 ; \frac{p_0 q_0}{n}\right)$

$q_0 = 1 - p_0$

$$Z = \frac{P - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \rightsquigarrow N(0,1) \quad z_1 = ? \; z_2 = ?$$

$P(z_1 < Z < z_2) = 1-\alpha.$

$$P\left(p - q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0 q_0}{n}} < p_0 < p + q_{1-\frac{\alpha}{2}} \sqrt{\frac{p_0 q_0}{n}}\right) = 1-\alpha.$$

On estimera p_0 sous le radical par p.

$$IC_\alpha(p_0) = \left[p \pm q_{1-\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}\right]$$

N. B.:

$IC_\alpha(p_0)$ est centré en p.

3/ Intervalle de confiance pour la variance:

$IC_\alpha(\sigma^2)$.

$a = ?$ $b = ?$ $P(a < \sigma^2 < b) = 1-\alpha$

On doit supposer la normalité de la pop.

a/ m connu :

Rappel: Théorème:

Soit $(X_i)_{i=1,2,\dots,n}$ v.a. ii.d.

$X_i \rightsquigarrow N(m, \sigma^2)$.

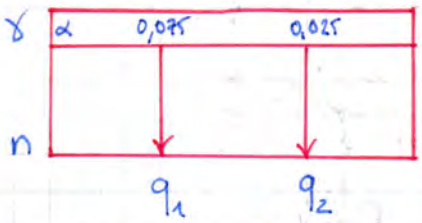
$$\Rightarrow Q = \sum_{i=1}^{n} \left(\frac{X_i - m}{\sigma}\right)^2 \rightsquigarrow \chi^2_n$$

$q_1 = ?$ $q_2 = ?$ $P(q_1 < Q < q_2) = 1-\alpha$

q_2 : $P(Q > q_2) = \frac{\alpha}{2}$

q_1 : $P(Q > q_1) = 1 - \frac{\alpha}{2}$.

q_1 : $\chi^2_n\left(1 - \frac{\alpha}{2}\right)$ $q_2 = \chi^2_n\left(\frac{\alpha}{2}\right)$.



$$P\left(q_1 < \frac{\sum (X_i - m)^2}{\sigma^2} < q_2\right) = 1-\alpha.$$

$$P\left(\frac{1}{q_2} < \frac{\sigma^2}{\sum (X_i - m)^2} < \frac{1}{q_1}\right) = 1-\alpha$$

$$P\left[\frac{\sum (X_i - m)^2}{q_2} < \sigma^2 < \frac{\sum (X_i - m)^2}{q_1}\right]$$

$$IC_\alpha(\sigma^2) = \left[\frac{\sum (X_i - m)^2}{\chi^2_n\left(\frac{\alpha}{2}\right)} ; \frac{\sum (X_i - m)^2}{\chi^2_n\left(1 - \frac{\alpha}{2}\right)}\right]$$

Lecture de $\chi^2_n\left(\frac{\alpha}{2}\right)$ et $\chi^2_n\left(1-\frac{\alpha}{2}\right)$

si ddl $= \gamma = n \leq 30 \rightarrow$ voir table

si $\gamma > 30 \Rightarrow \chi^2_n \simeq N(\gamma, 2\gamma)$.

$q_1 : P(Q > q_1) = \frac{\alpha}{2}$

$q_2 : P(Q > q_2) = 1 - \frac{\alpha}{2}$.

$Q \rightsquigarrow N(\gamma, 2\gamma)$.

b/ m inconnu :

m inconnu donc à estimer $\hat{m} = \bar{X}$

$$Q = \sum \left(\frac{X_i - \hat{m}}{\sigma}\right)^2 \rightsquigarrow \chi^2_{n-1}$$

$$Q = \sum \frac{(X_i - \bar{X})}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$$

$$IC_\alpha(m) = \left[\frac{\sum (X_i - \bar{X})^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)} ; \frac{\sum (X_i - \bar{X})^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}\right]$$

on a $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$

$\Rightarrow \sum (X_i - \bar{X})^2 = nS^2$

$$IC_\alpha(\sigma^2) = \left[\frac{nS^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)} ; \frac{nS^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}\right]$$

Calcul de $\chi^2_n\left(\frac{\alpha}{2}\right)$ et $\chi^2_n\left(1-\frac{\alpha}{2}\right)$

si $\gamma > 30$:

$\gamma > 30 \Rightarrow \chi^2_\gamma \simeq N(m, \sigma^2)$ $m = \gamma$

$\sigma^2 = 2\gamma$

$q_1 = ?$ $P(Q > q_1) = 1 - \frac{\alpha}{2}$.

$X \rightsquigarrow N(m, \sigma^2)$.

$\Downarrow$

$Z = \frac{X - m}{\sigma} \rightsquigarrow N(0,1)$

$P(X \leq x) = P(Z \leq z) = \Phi(z)$.

$z = \frac{x - m}{\sigma}$.

$P(X \geq x) = P(Z \geq z) = \Phi(-z) = 1 - \Phi(z)$

on a S^2

$P(Q > q_1) = \Phi(-z_\alpha) = 1 - \frac{\alpha}{2} = 0{,}975$

$-z_\alpha = 1{,}96 \Rightarrow z_1 = -1{,}96$. $(\alpha = 5\%)$

$q_1 = z_1 \sqrt{2\gamma} + \gamma$